Historic, archived document

Do not assume content reflects current scientific knowledge, policies, or practices.

ERROR PROFILE FOR MULTIPLE-FRAME SURVEYS

Norman D. Beller

U.S. Department of Agriculture

Economics, Statistics, and Cooperatives Service

ESCS-63

BIBLIOGRAPHIC DATA SHEET	1. Report No. ESCS-63	2.	3. Recipient's Accession No.
4. Title and Subtitle ERROR PROFILE FOR MULTIPLE-FRAME SURVEYS			5. Report Date August 1979
			6.
7. Author(s) Norman D. Beller			8. Performing Organization Rept. No. ESCS-63
9. Performing Organization Name and Address Statistical Research Division Economics, Statistics, and Cooperatives Service U.S. Department of Agriculture Washington, D.C. 20250			10. Project/Task/Work Unit No.
			11. Contract/Grant No.
12. Sponsoring Organization Name and Address			13. Type of Report & Period Covered Final
			14.

15. Supplementary Notes

16. Abstracts

Multiple-frame surveys are susceptible to errors that stem from associating the overlapping portions of the frames. Decreasing sampling error may improve precision, but it offers greater complications; nonsampling error would increase, leading to decreased accuracy and, perhaps, greater total error. This manuscript offers several procedures to improve consistency and accuracy of multiple-frame estimating by reducing nonsampling errors. Such procedures, when added to an optimum mix of sampling frames, would provide survey indications of greater reliability without increasing costs appreciably.

17. Key Words and Document Analysis. 17a. Descriptors

Errors
Estimates
Estimating
Sampling
Statistical analysis
Surveys

17b. Identifiers/Open-Ended Terms

Error profile
Estimating procedures
Frame surveys
Multiple-frame surveys
Nonsampling errors
Procedures
Sampling errors

17c. COSATI Field/Group 12-A, 12-B

18. Availability Statement	19. Security Class (This Report)	21. No. of Pages
Available from: NATIONAL TECHNICAL INFORMATION SERVICE, 5285 Port Royal Road, Springfield, Virginia 22161.	UNCLASSIFIED	
	20. Security Class (This Page) UNCLASSIFIED	22. Price

CONTENTS

ERROR PROFILE FOR MULTIPLE-FRAME SURVEYS

*Norman D. Beller**

INTRODUCTION

Ideally, a user of statistical data would have a measure of the total error arising from the statistical process; this rarely happens because samplers seldom, if ever, know the true value of the population from which the data are collected.

The total error generally comes from two sources. The first major source is the error that arises due to sampling---the process of measuring only a portion of a population and drawing inferences to the total population. The second source of error generally is referred to as nonsampling error and includes errors rising from an insufficient frame, a biased sampling procedure, the data collection procedure, the questionnaires, and the estimation procedure. Only a measure of the sampling error is provided in most sampling situations, rather than a measure of total error. Measures of nonsampling error components are not obtained in the bulk of the instances where sampling or a census is required, primarily due to the additional costs.

The preceding situation leads to a statistical paradox from a sampler's point of view. The total error is unknown and the sampler has at his or her disposal only the sampling error which may be somewhat controlled. One normally may assume that the primary purpose of sampling is to obtain needed information about the target population by measuring only a portion of the population due to costs, the destructive nature of sampling, or population characteristics changing rapidly over time. The information is to be obtained and estimated with a minimum sampling error, given appropriate cost restraints. Thus, a sampling statistician's goal in any survey, and probably moreover in a repetitive than in a single-time survey, is to minimize variation within cost restraints. Generally, the impact of the standard error minimization process results in a more complex survey design, questionnaire, and/or estimation procedure. These added complications can create situations that may increase nonsampling error. The paradox is that continued efforts to decrease sampling error (improve precision) often involve greater complications that increase the nonsampling error (decrease accuracy) which, in turn, may result in a greater total error.

Most error profiles will discuss the potential sources of nonsampling error. This paper concentrates on those errors where Economics, Statistics, and Cooperatives Service (ESCS) research has attempted either to identify or to measure a particular source of error. The bulk of the research effort has been directed to the multiple-frame hog and cattle surveys.

This report will show that there are several sources of nonsampling error in these surveys. These sources include failure to associate properly reporting unit and sampling unit, and failure to communicate clearly and concisely via the questionnaire, domain determination, estimation, and nonresponse.

---

*The author is chief of the Sample Survey Research Branch, with the Statistical Research Division of the Economics, Statistics, and Cooperatives Service, U.S. Department of Agriculture.

SAMPLING FRAMES

Two or more sampling frames are utilized in multiple-frame estimation. An area sampling frame is used with one or more list frames in the ESCS application. The area frame used by ESCS is all the land in the continental United States. All the land area has been classified by general land-use patterns. This classification is then used for stratification. The minimum stratification involves classifying land for high intensity agricultural use, medium intensity agricultural use, rangeland and woodland, cities, and towns. The sampling unit from the area frame is a given land area and is called a sample segment.

The area frame has the major advantages of being complete and current. It is generally easy to associate a sampling unit and a reporting unit. The major disadvantages are that it is not efficient for rare items that cannot be controlled by stratification, and is generally more costly than list frames in terms of data collection. The list frame for agriculture is a listing of names and addresses that are thought to be associated with agriculture.

In most applications, the sampling unit is a name and address, and the reporting unit is a farm that can be associated uniquely with a name and address. A major advantage of a list frame is that it is generally more efficient than an area frame, given one has adequate measures of size. A list frame with no or very poor measures of size provides minimal gains in efficiency over an area frame. Major disadvantages of a list frame are that it is rarely complete and deteriorates over time. It is also difficult to associate sampling and reporting units correctly. This difficulty frequently derives from an inadequate questionnaire or the data collection method used for part or all of the sample.

Unless one is uniquely able to make the necessary association so that the values of the characteristics for each sampling unit are accurately determined, unbiased results are not likely even though there is a random selection of sampling units.

AREA-FRAME ESTIMATION

Estimation from an area frame is a rather simple process. It expands the sampling unit (segment) totals by the reciprocal of the probability of selection. Ratio estimation is used infrequently because there is seldom a measure of the population mean corresponding to the variables that will succeed in reducing the overall variance. Also, ratio estimation is seldom used in a double sampling sense, not only because of the low correlations but because of the variance normally associated with the auxiliary variable. Double sampling to extend the base is an expensive procedure; consequently, the use of ratio estimation is limited.

Reported data must be associated with the sampling unit by a predescribed concept in all sample estimation procedures. This association can be achieved in three basic ways when dealing with the area frame estimates. The concept of the closed segment centers on the land area of the segment; thus, the reported data must be associated with the land area inside the segment. This association commonly is accomplished by accounting for all agricultural activities within the segment boundaries at a given time.

The closed segment concept is quite effective for the bulk of items collected in agricultural surveys that are highly associated with land area, such as acreage, cropland, and land use by specific types of crops. On the other hand, the closed segment may not be an appropriate concept for characteristics that must be associated with farms rather than with land area, such as the number of people who reside on farms.

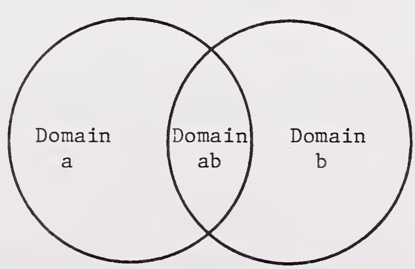
Another estimator has been developed based on applying the closed segment concept to residences. This estimator is generally called the open-segment approach. The open-segment concept associates the farm uniquely with a residence. If the farm residence is located within the boundaries of the segment, then the entire farm is associated with that segment; hence, the headquarters rule.

The open-segment estimator, using the headquarters rule, generally has a larger variance than estimators obtained using the closed approach. If the farm is located in more than one sampling unit, theoretically its probability of selection in an area sample depends upon the number of sampling units over which it extends. In application, the number of sampling units over which a farm extends is unknown and generally impractical to determine. There are, however, estimation procedures that may be used to reduce the variance of items that must be associated with a farm while using the open-segment concept. One such procedure associates with each segment that fraction of the total farm contained within the segment boundaries when it is located in more than one segment. The fraction of the farm associated with the segment is proportional to the acreage of the farm located within each segment. This estimator (weighted segment) is unbiased, given that the expected values of the two land variables for a particular farm are measured without error. If the land variables are not measured without error, either the variance is understated or the estimator may no longer be unbiased.

MULTIPLE-FRAME ESTIMATION

Multiple-frame estimation, again, implies the use of two or more sampling frames. The procedure allows greater coverage of the target population if no single complete frame exists. The procedure may also create a substantial amount of duplication of the target population between frames. Multiple-frame estimation provides greater efficiency if one can use less expensive data collection procedures on at least one of the frames. ESCS surveys generally use an incomplete list frame in combination with a complete area frame. The major objective for this multiple-frame estimation is efficiency, or a lower sampling error for a given level of cost. The following figure depicts the process of the theoretical development of multiple-frame sampling by Hartley (13). 1/

Figure 1: Two overlapping frames



Frame A　　　　Frame B

1/ Underscored numbers in parentheses refer to literature listed in the references at the end of this report.

Consider frame A the area frame; thus, domain b is the null or empty set by definition. Given simple random samples size n_a from frame A and n_b from frame B, Hartley's estimator is as follows:

$$(1) \qquad \hat{y}_H = \frac{N_A}{n_a}(y_a + p\, y_{ab}) + \frac{N_B}{n_b}(y_b + q\, y_{ba}),$$

with p and q constants such that p + q = 1. Since domain b is the null set, however:

$$(2) \qquad \hat{y}_H = \frac{N_a}{n_a}(y_a + p\, y_{ab}) + \frac{N_B}{n_b}(q\, y_{ba})$$

rewriting:

$$(3) \qquad \hat{y}_H = \frac{N_a}{n_a} y_a + p \frac{N_a}{n_a}(y_{ab}) + q \frac{N_B}{n_b}(y_{ba})$$

where y_a is a sample total obtained from those units in domain a. Domain a is called the nonoverlap domain and represents those units in the area frame that are not contained on the list frame. The area estimate (y_{ab}) from that portion of the target population covered by the list frame is called the area overlap domain. The estimated total (y_{ba}) is from the list frame for domain ab; it provides a second independent estimate of those elements common to both frames. Note that y_{ab} and y_{ba} are estimates for the same portion of the target population.

By setting p = 0 and q = 1, the following estimator, generally called a screening estimator, is obtained:

$$\hat{y}_H = \frac{N_a}{n_a} y_a + q \frac{N_B}{n_b}(y_{ba}). \qquad (4)$$

The screening estimator commonly is used when the value of p is expected to be quite small due to the costs of resulting variances being large in domain ab.

THE SURVEY DESIGN

The survey design for the multiple-frame estimator for hogs and cattle relies heavily on the ESCS June Enumerative Survey which is based upon the area-frame and conducted in late May. The area frame has been stratified by land use prior to sampling and has a total sample size of approximately 16,000 segments. The June Enumerative Survey provides estimates for major items at the State level with coefficients of variation from 3 to 12 percent. Large deviations from the segment mean have substantial impact on the sampling error in actual practice.

Lists of the largest operators (extreme operator lists) of both cattle and hogs have been developed and used in conjunction with the June Enumerative Survey in an effort to reduce the sampling error. Thus, no true area-frame estimates generated for hogs and cattle exist because the use of the extreme operator lists (lists of the largest producers) generates a multiple-frame estimator. However, only an area-frame estimate is generated for the balance of the survey items. Each State in the multiple-frame program of estimates has developed a list of farm operators other than extreme operators with control data for the number of hogs and cattle.

Generally, an attempt is made to have the list frames cover as much of the population of all farmers as possible. The list frame is formed into 5 to 10 strata for each species based upon a control variable. The sample size is approximately 1,900 farm operators per species for each State. The list of operators found in the area-frame sample is name-matched against the entire list frame after the June Enumerative Survey has been completed. The area-frame respondents that matched are classified as overlap (domain ab), and those that did not match are classified as nonoverlap (domain a). The multiple-frame hog survey is conducted four times a year to provide estimates as of December 1, March 1, June 1, and September 1. The multiple-frame cattle survey is conducted biannually to provide estimates as of January 1 and July 1. The screening estimator (equation 4) is used for all estimates.

SOURCES OF NONSAMPLING ERRORS

Nonsampling errors are difficult to measure because of the number of different kinds of error and the frequency at which a particular source of error occurs (a single source of error may take on the attributes of a rare item). In many instances, the sample size for a research study would have to be larger than the operational sample in order to obtain statistics that are significantly different. Such large sample sizes for research purposes are not practical in most cases. Therefore, the bulk of the studies have found differences which are not statistically significant. However, this report will include those differences whether they tested significantly different or not. Statistically significant differences have been so designated.

The nonsampling errors can have either a positive or a negative effect upon the estimator and, as a result, may have a balancing or compensating effect. One must proceed carefully in implementing changes; if the compensating nature of the errors is changed, an estimate with greater bias than before making the change may be obtained.

Multiple-frame surveys use two or more sampling frames. A multiple-frame estimator usually has more potential for nonsampling error than a single-frame estimator. A resulting estimator generally will have the nonsampling error, peculiar to each of the frames, as well as errors that may result in combining estimates for two or more frames. The sum of nonsampling errors could have a net effect less than any of the single frames due to balancing, or, in total, they could have a greater error. Nonsampling errors can arise from the area frame (y_a), from the list frame (y_{ba}), and from the overlap domain (y_{ab}) (shown in equation 3).

Questionnaire and Survey Concepts

The wording of questions to obtain needed information and ensure that the respondent understands the prevailing survey concepts has always been a matter of great concern. Howard T. Hovde sampled a group of experts in 1936 to find out what they considered the principle defects of research (16). The experts' most frequently mentioned criticisms were:

Improperly worded questionnaires	74 percent
Faulty interpretation	58 percent
Inadequacy of samples	52 percent
Improper statistical methods	44 percent

S. A. Stouffer arrived at nearly the same conclusions in 1950 (20). He found that error or bias attributed to sampling methods and questionnaire administration were relatively small when compared with errors attributed to different ways of wording questions---prompting the query, "If questionnaire wording is so important, why hasn't a questionnaire preparer done more to advance his phase of research?" One investigator suggested that a questionnaire preparer just doesn't exist---at least as a specialist: "The statistician is the only one among us who has a specialty. All the rest of the work comes under the jurisdiction of a jack-of-all trades. This man's job is to develop a questionnaire, pretest and revise it, have it printed, select, train and supervise the interviewers conducting a survey, analyze the results, write the report and present his findings (19)."

One might ask why question wording is so important. A questionnaire has several concepts to develop in addition to the data requirements for multiple-frame methodology. A questionnaire must provide information that will allow proper association of reporting and sampling units, overlap and nonoverlap determination, weights for the weighted estimator, and the basic information for the item of interest. The sample unit from the list domain is normally a name and address from the list, while the reporting unit from both the area and list frame is the operated land and the livestock on that land at the time of the interview. The association is established when the respondent is asked by phone, mail, or personal interview to report land owned, rented, or managed and land rented or leased to others. Land questions may be asked in a slightly different manner, depending upon the frame from which the respondent is obtained. Once the land area of the operating unit has been defined, the respondent is asked to report the total number of cattle and calves (hogs and pigs) on that land, regardless of ownership.

A 1974 study by William F. Kelly found interviewers considered the section of the questionnaire on acres owned to be one of the most difficult sections to complete and needed the most explanation to respondents (18). That study indicated that a fourth of the enumerators did not read the questions as printed on the questionnaire.

A study conducted the same year by Fred Vogel in Wyoming indicated that questionnaires obtained by mail required editing more frequently than those obtained by other methods (24). Bosecker and Kelly made a number of observations in a 1975 study conducted in Nebraska (3). They noted that the respondent's natural inclination was to report livestock on an ownership basis with no regard to where the livestock were located. Many of the errors found in that study were associated with cattle on a fee-per-head basis. They also noted that respondents failed to make the connection between the acres reported and the number of head and livestock regardless of the placement of the land questions. Bosecker and Kelly found the interviewer was the deciding factor on whether the respondent consistently adhered to the survey concept. A 1975 study in Kansas by Barry Ford tested the impact of establishing the livestock inventory prior to asking land questions (11). Ford found that moving the land questions to the end of the questionnaire increased the response rate, but failed to make a test of significance in livestock inventory because the power of the test was too low due to an inadequate sample size. He cautioned, however, that the true value calculated from the data was high enough to be alarming.

All the aforementioned studies point to the difficulty of using questions relating to land for the purpose of associating the reporting unit and the sampling unit. At best, this can have a very serious impact on the resulting estimate, since a failure of properly associating the reporting and sampling units will create nonsampling errors that affect the resulting estimate to the extent they are not self-balancing.

For another example of the use of the questionnaire to establish survey concepts, consider the calf-crop question that was dealt with in the Wyoming study by Vogel (24). The cattle multiple-frame estimate of calf crop is developed from two different reporting units. The reporting unit for the expected calf crop is cows and heifers which are expected to calve before December 31 on the operated land. Meantime, the reporting unit for calves already born is all calves born since January 1 on the land now operated. Considerable editing was required on the Wyoming questionnaires.

Nonsampling error could be created when each questionnaire is edited for consistency between sections when all data in a single questionnaire do not need to be consistent due to the use of two different reporting units. Table 1 shows the impact of editing for consistency on the survey estimates. The amount of editing on some questions resulted in changing the level of cattle and calves by an amount two or three times greater than the error caused by sampling. This amount of editing is cause for alarm in that it clearly shows a breakdown in the survey process.

A study in Ohio and Wisconsin by Hill and Rockwell further attempted to investigate survey concepts and the association of the reporting and sampling units (14). A test questionnaire was developed which essentially utilized more detail in screen-

Table 1--Effect of editing actions on survey estimates, Wyoming cattle and calf multiple-frame survey, July 1974 1/

Livestock and questionnaire version	Percentage change in estimates resulting from edit 2/	Relative sampling errors of final data
	Percent	
Calves born and still on ranch:		
Operational	+3.8	4.2
Text	-.5	3.3
Average	+1.5	2.6
Total calves born:		
Operational	+2.5	4.1
Test	+2.3	3.4
Average	+2.4	2.6
Cows and heifers expected to calve:		
Operational	-32.8	13.2
Test	-28.2	10.4
Average	-28.0	8.4
Calves weighing less than 500 pounds:		
Operational	+10.1	4.0
Test	+10.9	3.4
Average	+10.5	2.6
Total cattle and calves:		
Operational	+6.3	3.6
Test	+8.6	3.1
Average	+7.5	2.4

1/ Does not include data from extreme operators or the nonoverlap domain.
2/ Percentage change = edited value ÷ original value.

ing questions. The Ohio test questionnaire produced significantly higher estimates (20-percent increase in total hog inventory), while the results differed little in Wisconsin. The authors noted that the completion rate for the test questionnaire substantially differed between States---70 percent for the test questionnaire versus 90 percent for the operational questionnaire in Wisconsin, while both versions neared 80 percent in Ohio. The possible reason for the difference in the results points to respondents who received the operational questionnaire in Ohio and previously had been contacted twice producing a conditioning effect.

The same study attempted to determine the net effect of editing to make data conform to survey concepts. A second edit or review was completed and a resulting estimate comparing the second edit to another questionnaire obtained on a reinterview caused a reduction of approximately 6 percent of the estimates for both the operational and test questionnaires in both States. The proration of the partnership data serves as the primary reason for the edit changes.

Domain Determination

One of the most critical procedures in multiple-frame estimation is domain determination. Since the area frame is a complete frame, the overlap between the two frames is identified by determining whether each reporting unit found in the area sample could also have been selected from the list frame. Again, the sampling unit for area frame is a piece of land, and a name and address for the list frame. Since one cannot match pieces of land with names, it is necessary to associate a name and address with the land for a reporting unit for each sampling unit (segment). Overlap between the two frames is then determined by matching names associated with their respective reporting unit. This becomes extremely difficult with joint farming operations. The use of nicknames, nonperson names, names primarily generated for legal purposes, and minimal address information all add to the difficulties of matching accurately via the use of names and addresses.

Some of the earliest studies noted difficulties with domain determination. A 1965 Mississippi study evaluated Agricultural Stabilization and Conservation Service lists as a sampling frame, and investigated use of multiple-frame surveys to obtain unbiased estimates for crops (23). The report stated, "Results of this study indicate that many of the list units enumerated in the area sample could not be identified and led to substantial biases in the multiple-frame estimates. This bias may be larger than the reduction in variance realized for multiple-frame estimates when compared with direct-expansion estimates from the area sample."

List frames are updated once or twice a year. The area frame is required to estimate the incompleteness of the list frames; hence, the two frames must be kept independent. Knowledge of the existence of the unit in the area sample cannot be used to update the list. The name-matching procedure must be without error, and the frames must be kept independent for the estimates to remain unbiased during the process of domain determination. Resulting estimates will be biased to the extent that either of these factors fail.

The loss of independence is, perhaps, one of the most difficult biases to control. This loss is probably caused by each field office being responsible for conducting the surveys, determining domain, and updating lists. Office personnel spend a substantial amount of time working with lists and the area samples; therefore, the domain-determination procedures are somewhat subjective; with daily knowledge of list frames and area samples, independence is almost impossible to maintain. Past studies have indicated that the frames are not kept independent. The ratio of list overlap estimate for domain ab to area overlap for the same domain reaches 105 to 110 percent, when multiple-frame estimation is first instituted. The ratio decreases to 85 to 90 percent

after about 3 years of operation in a particular State or group of States. The downward movement of this ratio would lead one to suspect nonindependence.

The nonoverlap classification was observed by the length of time area-sampling units had been in use in a 1975 report (10). The area-frame sample was replicated and the replications utilized in an annual rotation program in those States. This procedure of sampling permitted study of the effect of time on domain determination. The percentage of tracts classified as nonoverlap declined significantly the longer the samples remained in area frame without rotation:

State	Number of years in sample	Percentage of tracts classified as nonoverlap	Standard error of percentage
	Number	------------- Percent	-------------
Nebraska	1	20.7	0.5
	2	14.8	.5
	3	12.7	.2
Missouri	1	22.0	.7
	2	18.3	.5

There was a dramatic downward trend in the nonoverlap estimates of hogs and cattle in Nebraska and hogs in Missouri. The standard errors of the nonoverlap estimates became so large, however, that the test of significance had no power. The report stated, "Obviously, the rotation group effect in nonoverlap percentages is a serious matter. It is not the application of the area-frame methodology that is called into question, but the application of the multiple-frame methodology. Was the list frame changed because of information from the area-frame sample, or was information accumulated over the years that indicated the correct nonoverlap classification? In either case, there is a problem with the nonoverlap classification procedures."

Another analysis explored the June 1973 Hog Survey (17). The list sample comprised over 95 percent of the population in that study. The analyst indicated that there could be a problem with domain determination: "The analysis of the June survey produced evidence that the area sample and the list sample were not estimating the same quantity. Such a situation could arise because of differing field procedures or because of errors in constructing the list or in identifying the overlap domain."

Further evidence that nonsampling errors are prevalent in domain determination may be gleaned from a 1974 report by Vogel and Bosecker (25). The following excerpts point out some of the detected errors that arose from operational procedures:

1. The name, originally coded overlap, was sometimes difficult to find on reexamination of the list. Nearly every State identified some errors resulting in additional nonoverlap tracts.

2. After a set of tracts was determined to be nonoverlap, some were not processed due to various reasons, mainly oversights.

3. The sampling frame used to identify nonoverlap was different from the sample frame from which the list sample was selected. For example, the alpha printout of the list frame contained names that did not have a chance to be selected by the sample select program.

4. Data were included for extreme operators in the area frame which should have been edited out.

5. Nonsampling errors detected in this analysis reduced the difference in levels of the multiple-frame and area-frame estimates by lowering the area-frame estimates and raising the multiple-frame estimates.

A 1977 report noted problems of joint operations when examined by a second edit determination (14). The purpose of the second edit was to ensure all concepts and overlap procedures had been followed correctly. The second edit results were then compared to reinterview questionnaires. The study found that 60 percent of all differences involved partnership arrangements. The major problem was determining if a partnership really existed or if it was an individually operated business. Depending on the determination of the actual tenureship, the current partial nonoverlap procedure may be seriously affected (the partial nonoverlap procedure relies on prorating the data to the nonoverlap and the overlap domains based on the number of chances the units had of being selected on the list frame). The differences found due to partnerships are shown in figure 2.

Figure 2: Summary of differences due to partnerships regardless of State or questionnaire version

Number of differences	Reasons
17	Second-edit interpretation was individual operation; reinterview interpretation was father-son partnership.
13	Second-edit interpretation was father-son partnership; reinterview interpretation was individual operation.
8	Second-edit interpretation was a partnership other than a father-son partnership; reinterview interpretation was individual operation.
5	Second-edit interpretation was individual operation; reinterview interpretation was a partnership other than a father-son partnership.
3	Selected combination of individuals does not operate land.
2	Change in number of partners from 2 to more than 2.
48	Total

The differences found in that study involved with nonpartnerships centered on the survey concept of obtaining livestock on the operated land regardless of ownership. Those differences follow:

Figure 3: Summary of differences due to nonpartnerships regardless of State or questionnaire version

Number of differences	Reasons
7	Failed to report hogs owned by someone else on his operated acres.
7	Additional hogs reported that were owned; reason hogs omitted from original report is unknown.
5	Included land rented out; hogs were on this land.
3	Reported breeding hogs, but left out feeder pigs.
3	Reason for difference is unknown.
2	Some hogs were temporarily on the father's operation, but all reported originally.
27	Total

A study conducted in 1976 evaluated alternative domain-determination methods with mounting evidence that nonsampling errors were prevalent there. Three different methods of domain determination were compared to the current partial nonoverlap procedure, which was implemented for the December 1971 Multiple-Frame Survey. The primary purpose of implementing the partial nonoverlap procedure was to minimize the effect of partnership and corporate farm operations on resulting sampling errors.

Again, the partial nonoverlap procedure relies on prorating the data to the overlap and nonoverlap domains based on the number of chances the units have had of being selected on the list frame. The methods differ only in the manner by which a name is associated with a unit of land. Specifically, variations in the four tested methods dealt mainly with the handling of joint operations; this was not duplicated on the list because all procedures were essentially the same for the name of an individual operator. Alternatives IIA and IIB differed only slightly from each other. They differed to a greater extent from the current procedure and required no proration of data to different domains. Alternative III eliminated proration of joint operation data. All partnership or corporate data were represented entirely by a single list frame sampling unit (name), or the partnership was edited entirely to the nonoverlap domain. A more detailed description of the four procedures may be found in the report, "An Evaluation of Alternative Methods of Overlap Determination (26)."

The study results are presented in table 2 as a percentage of the current procedure (partial nonoverlap). Analysis of the tabular data shows that substantial differences in the resulting estimates may be brought about by changing procedures. Use of a different procedure changed the level of the estimates as much as 5 to 6 percent in some States. Procedure II seems to have been the simplest procedure because it required the fewest assumptions and the least amount of knowledge about the joint operation. It was also the procedure mostly used prior to to December 1971 when the current partial nonoverlap procedure was instituted. The relative sampling errors show the partial nonoverlap procedure did not reduce the sampling error (comparisons of current versus alternative II).

Table 2--Multiple-frame survey cattle and hog estimates based on alternative nonoverlap procedures as a percentage of estimates from the current procedure, June 1975 1/

Procedure	Illinois	Iowa	Kentucky	Idaho	Minn.	Ohio	Total
				Percent			
Cattle and calves:							
Current	100.0 (3.6)	100.0 (3.3)	100.0 (3.6)	100.0 (3.5)	100.0 (3.8)	100.0 (6.7)	100.0 (1.7)
Alt. IIA	100.7 (3.5)	99.3 (3.3)	100.5 (3.6)	100.8 (3.4)	98.1 (3.6)	100.3 (6.7)	99.7 (1.6)
Alt. IIB	101.1 (3.5)	98.9 (3.3)	100.3 (3.6)	101.1 (3.4)	98.2 (3.6)		99.6 (1.6)
Alt. III	99.3 (3.9)	101.3 (3.7)	100.7 (3.8)	99.3 (4.0)	100.7 (4.4)	101.2 (6.9)	100.6 (1.8)
Hogs and pigs:							
Current	100.0 (6.6)	100.0 (3.6)	100.0 (8.1)	-- --	100.0 (6.4)	100.0 (7.0)	100.0 (2.7)
Alt. IIA	101.1 (6.4)	98.5 (3.6)	104.0 (9.0)	-- --	102.3 (6.3)	98.4 (6.4)	99.6 (2.7)
Alt. IIB	100.4 (6.4)	98.6 (3.6)	103.7 (9.0)	-- --	102.4 (6.3)	-- --	99.8 (2.8)
Alt. III	98.5 (7.1)	105.3 (4.3)	94.1 (8.2)	-- --	98.1 (6.6)	99.6 (9.0)	101.9 (3.1)

-- = Not applicable.

1/ Survey estimates for the current procedure are after a detailed review. Relative sampling errors appear in parentheses.

The report concluded: "Comparisons of Multiple-Frame Survey estimates of total cattle and total hogs on a state by state basis are not inconsistent with the theoretical proposition that each alternative nonoverlap procedure will yield the same results. When the state estimates are added together, the similarities are even more striking. Therefore, the choice among the alternative procedures should be based on ease of data collection and degree of nonsampling error."

It appears there is ample evidence that nonsampling errors are associated with domain determination, a very critical part of multiple-frame estimation. Nonsampling errors arising from domain determination normally may be regarded as an addition to any of the nonsampling errors associated with either an area or a list frame. It is safe to assume that the magnitude of errors arising from domain determination are positively correlated with the proportion of the universe covered by the list frame.

Size of List and Proportion to Sample

The purpose of multiple-frame cattle and hog surveys is to use a list as an efficient means of obtaining a desired sampling error and an unbiased estimate for the population of interest. It is generally more efficient to use a list of at least the larger operations in multiple-frame sampling to achieve a given sampling error rather than increasing the sample size drawn from the area sampling frame without knowledge of the location of large operators. The question becomes "How much of the universe should you attempt to cover with a list sample (how large should domain ab in figure 1 be)?"

The agency first had a policy of making the list as complete as possible and sampling the entire list using the extreme operator units. Over time, a rule of thumb was developed that the list should cover 90 percent of the item of interest for the multiple-frame cattle and hog surveys.

Throughout the history of the multiple-frame program, the contribution to the sampling error and the resulting estimates attributable to the area nonoverlap (domain a in figure 1) have been larger than desirable based on this approach. The area nonoverlap contributed about 20 percent of the estimate and 60 percent of its variability. Little, if any, success has been achieved in reducing contributions of the area nonoverlap either in level or variability regardless of the amount of effort placed on improving the list frame. This phenomenon can only be explained by recognizing what is taking place in the area frame. The item of interest becomes an increasingly rare item in the nonoverlap domain of the area frame as the list is made more complete and sampled in its entirety. The area nonoverlap estimator becomes less efficient as the item becomes rarer. Thus, the net result of increased resources spent for list improvement coupled with sampling the resulting list in its entirety are largely negated by decreasing efficiency in the area nonoverlap domain.

Why the concern about the portion of the list frame that is sampled for multiple-frame purposes? The sample can be optimally allocated to the various domains based upon cost and variability with proper statistical techniques. However, nonsampling error may arise as a result of multiple-frame estimation. For multiple-frame estimation to be unbiased, both domain estimates and domain determination must be unbiased. Domain determination decides in which domain each sampling unit should be placed. Each improperly classified unit will contribute to bias of the estimate.

Starting in 1974, a series of studies were conducted to determine the optimum mix of area and list frames (the optimum size of domain ab). The analyses sought to determine if the size of domain ab could be reduced without seriously affecting the sampling error, and thereby reduce the impact of nonsampling errors associated with domain determination.

A 1974 project provided results of an analysis of the 1973 Nebraska June Survey data (1). The analysis compared the respondents from the area-frame sample to those from the list frame. The list-frame strata codes were placed on the area-frame record for those whose names matched; summarization was then completed sequentially by dropping list-frame strata and enlarging area-frame nonoverlap. The study showed that by not sampling the strata with unknown, zero, or a small number for control from the list frame, the relative sampling error for the hog estimates would increase from 4.2 to 4.3 percent. The list universe and sample sizes would decrease from 54,193 to 24,877 and 1,748 to 1,286, respectively. The authors noted that the reduction in the size of the list frame would allow more time for duplication removal, identification and handling of joint arrangements, identification of overlap tracts, and detection of nonsampling errors. They specifically noted types of nonsampling errors, and that the magnitude of the nonsampling error would be reduced by sampling a smaller portion of the list frame. These nonsampling errors will be discussed elsewhere in this report.

Subsequently, the analysis was enlarged to four additional States for hogs and eight States for cattle and published in May 1974 (25). The results of the expanded analyses were quite similar to those found earlier in Nebraska. The impact on the relative sampling error for the universe and sample sizes show:

Item	Cattle			Hogs		
	Relative sampling error	Population	Sample size	Relative sampling error	Population	Sample size
	Percent	-----Number-----		Percent	-----Number-----	
Entire list	1.8	498,000	12,601	3.0	421,094	8,660
Zero and small-size strata deleted	1.9	78,000	5,538	3.6	78,015	4,375

A major conclusion of that analysis states "the relatively small decrease in sampling error obtained in the multiple-frame estimate by allocating 50 percent of the hog sample and 56 percent of the cattle sample to the zero or small-size group list strata is not providing a better estimate to the extent expected from the increased sample size."

The two preceding analyses were conducted using 1973 data. The analysis proceeded again on 1974 data to determine if the results would be consistent. This analysis included 12 States for cattle and four States for hogs (27). Again, the list frame and its resulting sample could be rather sharply reduced without substantially affecting the resulting estimate and its variance. The researchers proposed reducing the frame by varying degrees in the results. They proposed modified A and B procedures. Differences between the modified A and B procedures became the strata to be deleted from the list frame. Modified A was the most conservative approach, and consisted generally of dropping only the strata with a zero or unknown classification, while B extended the list strata to be dropped to those classified as having a positive but relatively small number of cattle or hogs. Results of that analysis show:

Item	Cattle			Hogs		
	Relative sampling error	Population	Sample size	Relative sampling error	Population	Sample size
	Percent	-----Number-----		Percent	-----Number-----	
Entire list	1.28	836,766	20,838	3.6	472,412	7,271
Reduced--Modified A	1.23	522,597	16,800	3.8	165,108	3,854
Reduced--Modified B	1.31	219,670	12,011	4.6	36,021	2,182

The report made the following recommendations: (1) the entire list frame should not be sampled for a given species, (2) in several States, even strata with small control should not be sampled, (3) a more efficient (costs versus relative sampling error) multiple-frame estimate will be obtained if smaller portions of the list frame are used for sampling purposes, (4) the levels of the estimates are not affected by sampling smaller portions of the list frame, and (5) the quality of the list frame needs to be improved considerably to achieve gains over the area-frame sample.

The area-frame survey is conducted in June and December. Thus, the area-frame data contribution to the multiple-frame program can be considered "free" for the June and December hog estimates and the January and July cattle estimates because the information would be collected in the enumerative survey regardless of a multiple-frame program. For the March and September hog surveys, a larger sample of the area nonoverlap would be required if the reduced list concept were made operational. A USDA study set out to determine the impact of the reduced list concept for the annual series of estimates, since the earlier analysis only considered the June cattle and hog surveys (27).

An additional sample of 200 nonoverlap tracts was selected to replace the zero and unknown strata dropped from the list frame for the March and September surveys. The study was conducted in four States for both cattle and hogs. The basic results of the study follow in table 3.

Table 3--Four-State study: Reduced list compared to current procedures for annual series of estimates

Item	Relative sampling error	Population	Sample size
	Percent	Number	Number
Cattle:			
Current procedure			
July	1.8	343,563	6,598
January	1.7	340,765	6,783
Reduced list			
July	1.7	177,838	4,977
January	1.8	186,873	5,063
Hogs:			
Current procedure			
June and September	2.6, 2.8	343,563	7,083
December and March	3.2, 3.4	340,765	7,434
Reduced list			
June and September	3.1, 3.6	65,599	4,537
December and March	3.2, 3.4	55,961	4,971

The study followed previous analyses and generally showed that the reduced list concept does not affect materially the resulting sampling errors. The preceding data display little if any change in precision except in June and September hogs. The reduced list concept produced a higher relative sampling error for these two surveys. Upon examining the individual State summaries, apparently the increase was caused in one State. The data indicate that a large respondence was in the area-frame survey for the zero list strata for two quarters, and is a reflection on the quality of control data used for stratification purposes. This large respondence could have shown up in the list sample as easily as in the area sample. The impact upon the resulting relative standard error would have been the same had it been in the list sample due to the probabilities of selection.

Thus, in the set of six surveys covered by the preceding table, 14,347 list-sample units were deleted, and 1,600 nonoverlap tracts added. This change in sample size did not change the sampling error of the estimate except in the large report just discussed.

Analyses all reached the same conclusion over several years for many different States. It is not necessary to sample the entire list frame for the multiple-frame cattle and hog program. Substantial reduction of sample size and burden may be realized, and the list frame may be substantially reduced. Nonsampling errors associated with domain determination may be minimized with this procedure.

Estimation

The screening estimator (equation 4) has been adopted for all multiple-frame estimators in ESCS. The screening estimator is obtained by adding an estimator for the area nonoverlap (list incompleteness) to the list estimator. There are several estimators for each of the components of the screening estimator. A direct expansion (expanding survey results by reciprocal of probability selection) may be applied to the open, closed, and weighted segment information to obtain the area-frame estimate. These methods of estimation were defined under area-frame estimation in the introduction. The open estimator seems the least efficient, while the weighted estimator is the most efficient. The list frame also uses the direct expansion estimator. The area frame is stratified by land use and the list frame is stratified by size of operation.

The current procedure is to use the weighted estimator for the area nonoverlap estimate (domain a). At the 1964 American Statistical Association meeting, Cochran compared the efficiency of the screening estimator (equation 4) with the estimator in equation 3 (4). Cochran developed cost functions which show that the choice of the estimators is related to the cost of collecting data in the respective frames. He stated, "On the average the screening estimator will have the lower variance whenever the cost of sampling from the supplementary frame is less than the difference between sampling from the 100 percent frame and screening members of the 100 percent frame in the supplementary frame." These principles are violated somewhat here. The area sampling frame and the major surveys based upon the frame (June and December Enumerative Surveys) are part of the ongoing program of estimates. Thus, the area-frame data are available and should be considered at a zero cost as an input to the multiple-frame program in June and December.

Another cost consideration is that, in most cases, a special list is developed for the multiple-frame program. Even if an existing list is available, the use of it in a multiple-frame program requires a higher standard of quality and more maintenance. Besides the cost of data collection, the additional costs of updating the list and maintenance for the list frame must be considered. Again, the area frame is supported by other resources. So when applying the cost considerations in choosing a multiple-frame estimator, the totality of costs becomes a factor. If this were done, the full

multiple-frame estimator would be utilized in place of or in addition to the screening estimator for the June and December multiple-frame surveys.

Why should there be concern about the choice of the screening or the full multiple-frame estimator in relation to nonsampling errors? To answer this question, one should also consider the errors arising from domain determination. Bias caused by improper domain determination is offset conceptually in the other frame. In other words, if the area-frame nonoverlap estimate is biased downwards by classifying certain area-frame respondents as overlap, when in truth they were not represented on the list frame, then the area overlap estimate would be biased upwards. Thus, the full multiple-frame estimator would reduce the impact of nonsampling errors in domain determination.

One of the reasons for choosing the screening estimator has been that the variance of the area overlap was of sufficient magnitude that any gains in efficiency from the use of the added information of the area frame have been largely negated. The nonoverlap portion of the estimate has contributed disproportionately to the variance of the resulting estimate. Recall the discussion of the nature of the variance of the nonoverlap under "Size of List and Proportion to Sample." The high variance of the of the nonoverlap estimate was caused by the sampling procedure which forced the occurrence of nonoverlap in the area frame to be a rare item by striving to have as complete a list frame as possible and sampling the entire list.

Researchers examined other estimators of the nonoverlap domain to obtain a more efficient multiple-frame estimator (2). They extended the full multiple-frame estimator to a stratum-by-stratum combination of estimates from two frames. Each area-frame respondent was coded according to the list-frame size group for each overlap respondent to obtain the estimator. Table 4 presents the results of this analysis.

Table 4--Multiple-frame livestock estimates using alternative estimators, June 1974

Multiple-frame estimator	Hogs			Cattle		
	Estimate	Standard error	Relative sampling error	Estimate	Standard error	Relative sampling error
	-----1,000 head-----		Percent	-----1,000 head-----		Percent
State A:						
$\hat{y}_A$ (area)	3,540.6	378.0	10.7	8,597.0	436.3	5.1
$\hat{y}_L$ (screening)	3,409.1	205.2	6.0	7,660.2	228.7	3.0
$\hat{y}_H$ (Hartley)	3,411.8	204.9	6.0	7,822.7	215.0	2.8
$\hat{y}_S$ (strata)	3,395.6	202.3	5.9	7,895.7	209.0	2.6
State B:						
$\hat{y}_A$ (area)	1,301.1	177.4	13.6	3,615.4	231.4	6.4
$\hat{y}_L$ (screening)	1,299.9	106.6	8.4	4,188.9	149.6	3.6
$\hat{y}_H$ (Hartley)	1,300.1	104.4	8.0	4,067.1	139.5	3.4
$\hat{y}_S$ (strata)	1,265.4	92.0	7.3	3,952.2	128.6	3.3

The analysis shows that the application of Hartley's estimator on either basis is more efficient. This is especially evident in State B because both the hog and cattle standard errors dropped lower than the screening estimator. Other studies have shown that additional gains in efficiency could be made in the nonoverlap estimate by utilizing the land-use stratification inherent in the area frame of the nonoverlap variance calculation (25). The report also noted that the major portion of the reduction realized in the strata estimator was obtained in the larger size livestock strata, and that only a minimal reduction occurred in the zero or smaller livestock strata. The reduced list concept could be utilized, and most of the gain from the strata estimator retained.

Again, the weighted estimator for the nonoverlap domain is utilized in the current program. The weights are based upon the proportion of the land area of the farm inside a segment to the land area of the entire farm. The condition required for the weighted estimate to be unbiased states that the sum of the weights equals one. However, a biased estimate results if the weight is not properly reported and/or calculated.

Generally, experience has shown that one of the more difficult reporting items for farms is the total land of the farming operation. Consequently, a December 1977 study investigated the use of the weighted segment estimator (15). A sample of the respondents was reinterviewed in three States for the December Enumerative Survey (DES) in an attempt to obtain better land data for weighting purposes. The report indicated difficulty with the weighted estimate and stated, "In all three States there was a significant downward bias in number of farm acres reported in the 1976 DES. This understatement of farm acres caused the weights (tract acres/farm acres) to be significantly too large. Therefore, even if the number of livestock was reported perfectly, the weighted livestock indications were subject to an upward bias." The extent of the bias, where the direct expansion of the reinterview data was compared with the original December expansions, appears in table 5.

Table 5--Reconciled data direct expansion as a percent of December Enumerative Survey expansion

Item	Indiana	North Carolina	Oklahoma	Total
		Percent		
Farm acres 1/	103	111	105	106
Farm cattle 1/	102	107	97	99
Farm hogs 1/	96	99	103	98
Tract hogs 2/	103	122	102	109

1/ Open segment estimator.
2/ Closed segment estimator.

The table presents results for farm acres, cattle, and hogs. However, of particular interest to the weighted estimate is the level of farm acres. The corrected farm acreage data ranged from 3 to 11 percent above the original survey indications. Since the weighted indications are obtained by the formula (acres in the segment ÷ total farm acres) x (farm livestock), one can observe that the bias of total farm acres results in a bias in the weighted livestock estimate. Thus, in this study, the weighted estimate has a built-in upward source of bias.

The differences in farm acreage were summarized by specific causes. The reasons for the difference were as follows:

Figure 4: Number of differences (DES versus Reconciled) in entire farm acres by reason

Reason	Three States combined
REPORTED FARM ACRES TOO LOW IN DES	
Acreage was estimated	26
Miscounted acreage, left some out	24
Entire parcel left out - idleland or woodland	19
Failed to report land rented from others	15
Failed to report land not in use	13
Attributed to a different respondent	11
Omitted entire farm acres	8
Split tract not picked up	7
Don't know	6
Misunderstood questions	5
Entire parcel left out - pasture	3
Failed to report land in a separate location	2
Left out operated land owned by family members	2
Didn't remember first interview	2
Land was to be sold in the near future	2
REPORTED FARM ACRES TOO HIGH IN DES	
Acreage was estimated	18
Included land rented out	15
Included public land	10
Attributed to a different respondent	8
Split tract not picked up	6
Miscounted acreage, included too much	5
Don't know	5
Included land operated by family members	5
Misunderstood questions	4
Included land in a different business arrangement	4
Included entire parcel of nonoperated idleland or woodland	3
Didn't remember first interview	2
Miscellaneous	2
Total	232

Reductions in the sampling error by using the weighted estimate for nonoverlap domain may be offset by nonsampling errors if the information used to calculate weights cannot be collected accurately.

There are several estimators available and each can serve a valuable function. The screening estimator has the advantage of eliminating the need to collect data from the area overlap domain (domain ab). The weighted estimate allows for telephone and mail data collection procedures, while the tract would require personal interview and thus be more expensive. Use of the full multiple-frame estimators will minimize the impact of nonsampling errors arising from domain determination. It also will provide a check on nonsampling errors associated with the screening and weighted estimators.

The following tabular array shows the types of estimation available without additional data collection:

Estimator 1/	Cattle		Hogs			
	Jan	July	Dec	March	June	Sept
Screening estimator	*	*	*	*	*	*
Hartley's estimator	*	*	*		*	
Stratified estimator	*	*	*		*	

1/ Each of these should be calculated on a weighted and tract basis with the exception of the March and September hog surveys, where only the weighted estimate is used for the screening estimator.

Nonresponse and Data Imputation

A concerted effort was initiated to research data imputation procedures for missing records in 1976. A missing record procedure may be thought of as an imputation or an estimation procedure. The missing record procedure may be applied to the estimator or to each questionnaire. The current procedure for missing records assumes the distribution of the item of interest for respondents is the same as for nonrespondents. This assumption works reasonably well when the distributions are the same or very similar. A bias in the resulting estimate occurs when they are not the same.

One would suspect the bias would be negative, because most experience and studies have shown that the nonrespondents are generally larger operators in terms of the items of interest than the respondent. One also might suspect that the proportion of the sample reporting a zero amount of the item of interest also would be smaller for the nonrespondents, since the participation rate for those having a zero amount of the item of interest is generally higher. Both of these factors contribute to the bias previously mentioned.

ESCS experience shows over the past several years that the nonresponse problem is greater in the list frame as opposed to the area frame. The area-frame nonresponse rate ranges between 2 and 10 percent, while the list frame is substantially larger. The nonresponse rate has been gradually increasing in recent years. Other efforts have been used in public relations and various survey procedures to counteract increasing nonresponse. These efforts, however, have not been successful to data in reversing the upward trend. The size of the control variable increases with the numerical designation of the strata (table 6). The data show that the nonresponse problem is greater in the stratum with larger control numbers.

A preliminary report titled "Missing Data Procedures: A Comparative Study" investigated six missing record procedures: a double-sampling ratio procedure; a double-sampling regression procedure; and four variants of a hot deck procedure (8). The analysis showed no significant differences in the resulting means from having used each of the six procedures. The research found that each of the procedures reduced the relative bias that would have resulted with the assumption that the mean of the respondents was equivalent to the mean of the nonrespondents; this reduction was made by more fully utilizing the control data. The reduction in relative bias among the procedures ranged from 8 to 26 percent. The large reduction in relative bias would have been achieved if an auxiliary variable with higher correlation were available.

Research continued on the missing record problem, and a sequel in June 1978 was published (9). The ratio, regression, and hot deck procedures again were analyzed. A

Table 6--Nonresponse rates by selected stratum for five Midwestern States, June 1978 multiple-frame hog survey

Stratum	1	2	3	4	5
			Percent		
1	10	17	15	13	6
2	14	8	26	12	7
3	25	23	34	27	13
4	28	23	33	38	20
5	23	28	34	39	22
6	29	32	22	39	26

balanced repeated replication design was integrated into each missing data procedure to compensate for the underestimate of the variance by the hot deck procedures found earlier. This procedure then provided unbiased estimates of the variance for all imputation procedures. A major conclusion shows that the auxiliary variables or control data were rather poorly correlated with the item of interest. Most of the livestock data show that the correlation with control, while it varies from State to State, is normally about 0.3 or less while analysis by the previous study indicated the correlation should be approximately 0.6 or more before any of the missing record procedures would make a significant improvement over the operational procedure of substituting the mean of the respondents for nonrespondents. The previous study recommended that if a missing record procedure were to be implemented under present conditions it should be a ratio procedure, a hot deck procedure using balanced repeated replications, or the hot deck procedure without replication.

Again, the need for obtaining better control data is noted in a 1978 working paper (7). The working paper made the following recommendations: (1) monitor the quality of control or auxiliary data, (2) examine methods used in constructing control variables, and (3) reevaluate the number of list strata.

The preceding analyses of various imputation procedures for missing records concluded that the current procedures could not be improved unless auxiliary or control variables were improved. Having reached this conclusion, additional work was directed at measuring and/or minimizing the downward relative bias caused by nonrespondents.

A study was developed to test the assumption that the mean of the nonrespondents differed from the mean of the respondents. The January and July 1977 Cattle Surveys in Colorado and the June and September Hog Surveys in Minnesota and Nebraska formed the basis for this study. The procedure identified the refusals and then employed specially selected enumerators in an all-out effort to convert these refusals during the survey period. Nonrespondent means could be estimated for those who only report in one of the surveys as a separate domain. The work was restricted to the list respondents in smaller size list strata.

The results appear in a 1978 study (12). Two major results highlight the study. The first is that the relative bias generally ranged from about 2 to 5 percent. Second, while the efforts to convert refusals from the previous survey were successful---ranging

from 40 to 50 percent converted---the overall refusal rate changed very little. This leads one to suspect that refusals are directly related to the quality of enumerators. The authors also noted, "Multiple-frame livestock estimates are generally biased downward because nonrespondent means on the list frame tend to be larger than respondent means."

A 1978 Nebraska study addressed the objective of improving the estimation of nonrespondents by obtaining additional information (6). An estimator was developed which would compensate for the greater proportion of positive reports in the nonrespondent group. Interviewers attempted to determine the presence or absence of hogs for the nonrespondents during the survey.

Using a weighted average of the proportion in each stratum, the researcher estimated operations in the population having hogs was 28 percent among respondents and 63 percent among nonrespondents. The resulting estimate moved the original list-frame estimate upwards by nearly 6 percent. A followup study showed a 2- to 6-percent downward bias by using the current procedures (5).

The researcher noted, "It would, therefore, seem more reasonable to identify positive nonrespondents in the sample so that their relative influence may at least be represented by the mean number of head reported by the respondents. This may still be only a partial adjustment if the mean number of head owned by nonrespondents should actually be higher than the respondents' level. However, the first step is feasible and is recommended for the operational program."

Finally, the research in nonresponse and data imputation has shown that there are feasible methods of reducing the relative bias caused by substituting respondent means for nonrespondent means. However, all viable procedures rely on high quality control data. The quality of control data must be improved before any improved imputational procedures may be adopted. An estimator has been developed for now that adjusts for the differing amounts of zero reports in the respondent and nonrespondent groups. Use of this estimator would reduce the relative bias and increase the list-frame estimate of the overlap domain.

CONCLUSION

In 1974 Her Tzai Huang stated, "The analysis of the June survey produced evidence that the area sample and list sample were not estimating the same quantity. Such a situation could arise because of differing field procedures or because of errors in constructing the list or in identifying the overlap domain (17)." This quotation perhaps best describes the ESCS experience with multiple-frame surveys. Research over the past several years validated his statement for each of the reasons cited. One might ask the question, "What can be done to minimize the impact of nonsampling error for the resulting estimators?" Adoption of the following general procedures may prove helpful:

A. Conduct proper testing before instituting new questionnaires or making major changes in questionnaires or survey procedures.

B. Build an adequate quality control program into the survey system so that constant monitoring can be accomplished.

C. Minimize procedures that are more susceptible to nonsampling errors.

D. Provide sufficient analysis data along with the estimators. This would lead to an extension of a quality control program.

E. Include sufficient consistency checks between items within a given questionnaire as well as between surveys when the respondent is the same.

F. Simplify procedures and/or estimators whenever possible. The complex survey procedure which does not reduce substantially the sampling error may have the potential for reducing the value of the resulting estimates.

Multiple-frame sampling subjects one to the errors associated with each of the frames plus those inherent to multiple-frame sampling. Errors inherent in multiple-frame estimation are mainly those arising from improper association of the reporting unit and sampling unit and those caused by domain determination.

Research studies point to difficulties in the current procedure of associating reporting and sampling units by use of land questions. The respondents' natural inclination is to report on an ownership basis. It is quite possible and, in fact probable, that the list frame is on an ownership basis while the area frame is on a land-operated basis. To the extent that this occurs, there is an upward bias in the resulting multiple-frame estimate. Perhaps, a breakdown of livestock on an ownership basis needs to be investigated for the area frame.

Studies also show that some of the concepts in the questionnaire are not clearly understood by respondents. These concepts should be simplified whenever possible. Not only would such simplifications improve the estimate, but they might also result in the respondents having a more positive attitude towards crop reporting.

One of the most critical procedures in multiple-frame estimation is domain determination. The development and use of list frames must be kept independent of the area frame. There is evidence that strict independence is not always maintained.

The procedures for domain determination are somewhat subjective, and the materials used in the process are less than adequate. Making the domain-determination procedure more objective would help reduce these errors. The partial nonoverlap procedure is a further complication of an already difficult problem and probably requires more information than is available in the bulk of the list frames utilized.

A relatively simple and straightforward procedure is to limit as far as practical the portion of the universe covered by the list frame to minimize the impact of nonsampling errors arising from domain determination. Ideally, the list frame would be limited to the point where the increase in sampling error is no greater than the decrease in nonsampling errors. Studies have shown that gains can be made by using the "reduced list" concept, even though this point cannot be exactly determined because of the inability to measure nonsampling errors. Several analyses have been made over the past several years which show that current procedures of maximizing the impact of the list frame have been ineffective in improving the accuracy of hog and cattle estimates. All analyses support the conclusion that the size of the list frame and the number of sampled strata could be reduced without measurably affecting the sampling error.

Data from the area frame required for the full multiple-frame estimator are available and in machine media for all but the March and September Hog Survey. Collecting data from the area frame for the sampled list strata would probably not be cost effective in March and September. Use of the screening estimator will maximize the errors associated with domain determination. However, comparison of the full multiple-frame estimator and the screening estimator will provide a rough measure of the nonsampling error arising from domain determination. The weighted estimator for area nonoverlap is subject to error caused by the difficulty respondents have in reporting total land in farm. It appears that more effort would be advisable in obtaining total land in farm

or development of a weight, other than total land. Use of the weighted estimate for the March and September Hog Surveys is probably required because of data collection costs. Nonsampling errors arising from poorly reported total land may be monitored by calculating all estimates on a weighted and tract basis. Further gains in efficiency may be obtained by the use of adding stratification to the multiple-frame estimator.

All estimates should be obtained since additional data need not be collected to utilize the multiple-frame stratified estimators for June and December hog estimates and January and July cattle estimates. This would provide the commodity experts with additional indications for their use in arriving at official estimates. The use of the various estimators would provide a means of analyzing nonsampling errors which come about from domain determination and may change from one year to the next. Further analysis of data should result in a sounder official estimate through the more effective use of data processing capabilities.

Nonresponse is large and growing. Normally we have been able to control nonresponse in the area frame; however, this is not the case in the list frame. Control data must be improved before data imputation procedures can become beneficial. Estimation procedures have been developed and should be utilized, which allow for a smaller number of zero reports in computing the nonrespondent mean.

REFERENCES

1. Bosecker, Raymond R., and Fred A. Vogel. "Analysis of 1973 Nebraska June Enumerative Survey and Multiple-Frame Survey Livestock Estimates," U.S. Dept. Agr., Stat. Rptg. Serv., Sept. 1974. (Unpublished report.)

2. __________, and Barry L. Ford. Multiple-Frame Estimation with Stratified Overlap Domain, U.S. Dept. Agr., Stat. Rptg. Serv., July 1976.

3. __________, and William F. Kelly. Summary of Results from Nebraska Survey Concept Study, U.S. Dept. Agr., Stat. Rptg. Serv., Nov. 1975.

4. Cochran, Robert S. Multiple-Frame Sample Surveys, proceedings at American Statistical Association meetings, Univ. Wyoming, Dec. 1964.

5. Crank, Keith N. The Use of Current Partial Information to Adjust for Nonrespondents, U.S. Dept. Agr., Econ. Stat. Coop. Serv., Apr. 1979.

6. __________. "Study of Nonrespondents in Nebraska March Hogs Survey, 1978," U.S. Dept. Agr., Econ. Stat. Coop. Serv. (Unpublished report.)

7. Ford, Barry L. A General Overview of the Missing Data Problem, U.S. Dept. Agr., Econ. Stat. Coop Serv., Aug. 1978.

8. __________. Missing Data Procedures: A Comparative Study, U.S. Dept. Agr., Stat. Rptg. Serv., Aug. 1976.

9. __________. Missing Data Procedures: A Comparative Study (Part 2), U.S. Dept. Agr., Econ. Stat. Coop. Serv., June 1978.

10. __________. Rotation Group Effects in SRS Surveys, U.S. Dept. Agr., Stat. Rptg. Serv., June 1975.

11. __________. The Effect of Land Questions on the Multiple-Frame Hog Survey, U.S. Dept. Agr., Stat. Rptg. Serv., June 1975.

12. Gleason, Chapman P., and Raymond R. Bosecker. The Effect of Refusals and Inaccessibles on List-Frame Estimates, U.S. Dept. Agr., Econ. Stat. Coop. Serv., Aug. 1978, p. 16.

13. Hartley, H. O. Multiple-Frame Surveys, Proceedings of the Social Statistics Section, proceedings at American Statistical Association Meetings, Minneapolis, Minnesota, Sept. 1962.

14. Hill, George W., and Dwight A. Rockwell. Associating a Reporting Unit with a List Frame Sampling Unit in Multiple-Frame Sampling--Ohio and Wisconsin, U.S. Dept. Agr., Stat. Rptg. Serv., June 1977.

15. __________. and Martha S. Farrar. Impact of Nonsampling Errors on Weighted Tract Survey Indications, U.S. Dept. Agr., Stat. Rptg. Serv., Dec. 1977.

16. Hovde, Howard T. "Recent Trends in the Development of Market Research," American Marketing Journal, no. 3, 1936, p. 3.

17. Huang, Her Tzai. "The Relative Efficiency of Some Two-Frame Estimators," Statistical Laboratory, Iowa State Univ., Mar. 1974, p. 27. (Unpublished report.)

18. Kelly, William F. Enumerator Evaluation of 1973 December Enumeration and Multiple-Frame Hog Surveys, U.S. Dept. Agr., Stat. Rptg. Serv., Mar. 1974.

19. Payne, Stanley L. The Art of Asking Questions, Princeton, N. J.: Princeton Univ. Press, 1951.

20. Stouffer, S.A., and others. Studies in Social Psychology in World War II, vol. IV, Princeton, N. J.: Princeton Univ. Press, 1950, p. 709.

21. U.S. Department of Agriculture. "Reduced Livestock List Study--Multiple-Frame Surveys, June 1977-March 1978," Econ. Stat. Coop. Serv., Oct. 1978. (Unpublished report.)

22. __________. Maintaining Statistics with Integrity, proceedings at National Conference of the Statistical Reporting Service, Dulles, Virginia, Feb. 1975, pp. 59-60.

23. __________. "1965 Mississippi Multiple-Frame Study," Stat. Rptg. Serv., Jan. 1965. (Unpublished report.)

24. Vogel, Fred A. How Questionnaire Design May Affect Survey Data--Wyoming Study, U.S. Dept. Agr., Stat. Rptg. Serv., Mar. 1974.

25. __________, and Raymond R. Bosecker. Multiple-Frame Livestock Surveys: A Comparison of Area and List Sampling, U.S. Dept. Agr., Stat. Rptg. Serv., May 1974.

26. __________, Raymond R. Bosecker, and Dwight Rockwell. Multiple-Frame Livestock Surveys: An Evaluation of Alternative Methods of Overlap Determination, U.S. Dept. Agr., Stat. Rptg. Serv., June 1976.

27. __________, and Jerry G. Thorson. Optimizing the Use of List and Area Samples in Livestock Multiple-Frame Surveys, U.S. Dept. Agr., Stat. Rptg. Serv., Sept. 1973.

*U.S. GOVERNMENT PRINTING OFFICE; 1979 280-931/ESCS-137

UNITED STATES DEPARTMENT OF AGRICULTURE
WASHINGTON, D.C. 20250

POSTAGE AND FEES PAID
U.S. DEPARTMENT OF
AGRICULTURE
AGR 101
THIRD CLASS

U.S.MAIL